DOCUMENT RESUME

ED 070 470 LI 004 022

AUTHOR	McGrath, William E.; Simon, Donald J.
TITLE	LNR: Numerical Register of Books in Louisiana Libraries; Basic Documents, I.
INSTITUTION	Louisiana Library Association, Baton Rouge.
SPONS AGENCY	Louisiana State Library, Baton Rouge.
PUB DATE	Dec 72
NOTE	55p.; (4 References)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Books; Computer Output Microfilm; Computers; Libraries; *Library Collections; Microfiche; *Union Catalogs
IDENTIFIERS	*Louisiana

ABSTRACT

This collection of documents contains the following: (1) Proposal to Implement a Plan to Create a Retrospective Union of Books Based on Library of Congress Card Numbers in Selected Lousiana Libraries Using a Computer to Sort and Print the numbers; (2) Proposal to Create a Retrospective Union Catalog of Books in Participating Louisiana Libraries Based on Library of Congress Card Numbers Using the Computer to Sort and List the Numbers in Printed Form; (3) A One Million Volume Computer Output Microfiche Numerical Union Catalog in Louisiana, with a Statistical Summary; and, (4) Predicting Title Multiplication (Overlap) in a Union Catalog of Sixteen Louisiana Libraries Using Regression Analysis. (Related materials are available as LI 004 023.) (SJ)

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED 070470

LNR: Numerical Register of Books in Louisiana Libraries

Basic Documents, I

William E. McGrath

Donald J. Simon

Compiled for the Louisiana Library Association

FILMED FROM BEST AVAILABLE COPY

LI 004 022

Lafayette,
Louisiana
December, 1972

CONTENTS

1. Proposal to Implement a Plan to Create a Retrospective Union Catalog of Books Based on Library of Congress Card Numbers in Selected Louisiana Libraries Using a Computer to Sort and Print the Numbers.

2. Proposal to Create a Retrospective Union Catalog of Books in Participating Louisiana Libraries Based on Library of Congress Card Numbers Using the Computer to Sort and List the Numbers in Printed Form.

3. A One Million Volume Computer Output Microfiche Numerical Union Catalog in Louisiana, with a Statistical Summary.

4. Predicting Title Multiplication (Overlap) in a Union Catalog of Sixteen Louisiana Libraries Using Regression Analysis.

PROPOSAL TO IMPLEMENT A PLAN TO CREATE A RETROSPECTIVE UNION CATALOG OF BOOKS BASED ON LIBRARY OF CONGRESS CARD NUMBERS IN SELECTED LOUISIANA LIBRARIES USING A COMPUTER TO SORT AND PRINT THE NUMBERS

PROPOSAL TO IMPLEMENT A PLAN TO CREATE A RETROSPECTIVE UNION CATALOG OF BOOKS BASED ON LIBRARY OF CONGRESS CARD NUMBERS IN SELECTED LOUISIANA LIBRARIES USING A COMPUTER TO SORT AND PRINT THE NUMBERS

Submitted to:	Louisiana State Library Under the LSCA Title III Program (Interlibrary Cooperation)
Submitted by:	University of Southwestern Louisiana for the LC Card Union Catalog Committee of the Louisiana Library Association
Project Director:	William E. McGrath, Director of Libraries University of Southwestern Louisiana
Systems Analyst:	Donald Simon University of Southwestern Louisiana
Telephone:	(318) 233-3850, Ext. 378
Date Submitted:	February 19, 1971
Amount Requested:	$8,000.00
Duration of Project:	From date of approval through June 30, 1971

I certify that the distribution of costs as shown in the proposal conforms to the usual accounting practices of the institution.

Richard C. Delcambre, B.A.
Business Manager

Clyde L. Rougeou, Ph.D.
President, University of Southwestern Louisiana

Ray Authement, Ph.D.
Vice President, University of Southwestern Louisiana

William E. McGrath, A.B., M.A.L.S.
Project Director

Jack D. Testerman, B.A., M.S.
Director of Institutional Research

PROPOSAL TO IMPLEMENT A PLAN TO CREATE A RETROSPECTIVE UNION CATALOG OF BOOKS BASED ON LIBRARY OF CONGRESS CARD NUMBERS IN SELECTED LOUISIANA LIBRARIES USING A COMPUTER TO SORT AND PRINT THE NUMBERS

Background

In 1969, Mr. William E. McGrath approached several academic libraries with a proposal to create a union catalog of books in selected Louisiana libraries. The catalog would consist of LC card numbers sorted and listed in the computer. In August 1969, he presented his idea to Miss Sallie Farrell, State Librarian, and Mr. John Richard, then President of the Louisiana Library Association. The idea was well received and a feasibility committee under the sponsorship of LLA was formed, chaired by Mr. William E. McGrath.

In the fall of 1970, with 57,630 volumes listed on the USL-based computer, the Committee submitted a proposal for $87,800 to the Council on Library Resources. The proposal presented to the Council, as given in the proposal's abstract, was as follows:

A. to create a retrospective union catalog of all those books represented by LC cards in major participating Louisiana libraries including academic, public and the state library;

B. to print the complete catalog, consisting of LC card numbers, and location symbols and distribute copies to participants, or to convert the record to microfilm directly from the computer;

C. to explore the practicality of permanent resident storage of the catalog of LC numbers on magnetic disk;

D. to determine the feasibility of making remote inquiry into the magnetic disk via teletypewriter (TWX) from each of the contributing libraries;

E. to determine whether combination of resident storage and print-out is practical;

F. to develop procedures for continuing participation after the completion of the retrospective catalog;

G. to analyze the patterns of acquisitions and overlap from the statistics generated by contributions and to determine whether predictions about overlap can be made;

H. to determine what new cooperative agreements concerning acquisitions and interlibrary loan should be made as a result of the project development and statistical findings.

It was well received by the Council, but they felt additional experience with the current catalog was needed. They indicated a willingness to receive another proposal at a future date.

The committee met in December 1970, to break the proposal into several parts with the thought that the parts of the proposal might be funded individually by one of more agencies.

It was generally agreed that the retrospective part of the proposal would lend the greatest application to usefulness.

Objectives

The objectives of this proposal for LSCA Title III funds are:

1. to implement, on a small scale, that part of the original proposal describing a retrospective Union Catalog
2. to provide easy access to the collections of the major academic libraries and the State Library for the purpose of interlibrary loan.
3. to make a computer printout of the Catalog available to all the original participants and to as many other libraries as possible--especially those served by the State Library. The goal will be to print 250,000 numbers. Fifty copies of the printout will be made, bound and distributed.

Utility

The catalog is used as follows. A librarian endeavoring to locate a

book records the LC card number at the time of verifying the author and title. The librarian then checks the LC card number against the list to ascertain the possible location of the title. The title would be requested by author and title from the library owning the book.

The titles will include the complete collection of the State Library having LC card numbers (about 100,000 titles), and about 150,000 from the combined collections of LSU-Baton Rouge, LSU-New Orleans, USL, and Louisiana Tech University.

During the past year, the University of Southwestern Louisiana's computer has been used to list the current acquisitions contributed by the Louisiana State Library, the New Orleans Public Library, and the following academic libraries: Louisiana Tech, University of Southwestern Louisiana, LSU-Baton Rouge, LSU Medical Center, Grambling, Centenary, LSU-Alexandria, Southeastern, Northwestern, Northeastern, Loyola, LSU-Shreveport, LSU-New Orleans and Nicholls State. About 70,000 volumes representing 56,000 titles have been listed so far in the current catalog. Even though use of this limited collection of titles can be expected to be considerably less than use of a full retrospective listing would be, there has been a demonstrable use for interlibrary loan. Interlibrary loan data for all the participating libraries is not yet available, so that no specific figures or percentages of requests found can be given at this time.

Use of the projected 250,000 volume retrospective catalog should be substantially greater than the 70,000 volume current catalog. Indeed, it will be possible to combine the two catalogs, making even greater use likely.

Methodology

During this phase of the project the five libraries represented by members of the Union Catalog Committee will attempt to extract and prepare for com-

puter input a substantial portion of their holdings of books with LC card numbers. The five libraries and their projected number of contributions are:

Louisiana State Library	100,000
USL	50,000
La. Tech	40,000
LSU-Baton Rouge	50,000
LSU-New Orleans	40,000

This totals to 280,000 numbers, but it should be understood that such a total can be no more than a rough projection because of the many factors involved such as libraries, people, books, equipment and supplies. The goal has therefore been set at 250,000.

The inclusion of these libraries does not necessarily exclude other libraries from participating in this phase of the project. Other libraries will be included--if their data can be prepared--but will not be funded from this budget. The ultimate goal, of course, is to include as many libraries and as many volumes as possible, therefore other funds will be sought from the Council on Library Resources or other sources.

The ultimate goal will still be the one-and-one-half to two million LC numbers projected in the original proposal.

Each of the several participating libraries will keypunch the LC card number for the contracted number of titles in its collection, or contract to have them done.

Criteria for Selection of Participating Libraries

The five libraries participating in this phase of the project are all but one of those represented on the committee. The pre-existence of this committee made possible a quick agreement in a complex situation. These five libraries are the largest of all the participants in the original project, and have also contributed the largest number of holdings to the current catalog. Another factor contributing to their agreement is the availability of keypunch equipment

or their willingness to assume keypunch responsibility for the other participants.

The five participating libraries are:

Louisiana State Library

University of Southwestern Louisiana

Louisiana Tech University

Louisiana State University - Baton Rouge

Louisiana State University - New Orleans

Budget

By far the greatest portion of the budget will be expended to support clerical and keypunching costs by the participating libraries. The clerical costs will vary widely from library to library. For example, at Louisiana State Library the particular form of their card catalog will make it much more difficult to extract LC card numbers than at many of the other libraries. Hence LSL's clerical costs will be substantially higher than in the other libraries. It should be noted, however, that there is substantial justification for including the State Library. The State Library has no restrictions whatsoever on loaning any of its books to any other library in the State; and loaning to other libraries is one of its primary services. In fact, it might be suggested that no cost should be spared in listing the State Library's holdings for the benefit of all the other libraries.

For internal reasons, two libraries, LSU-Baton Rouge and LSU-New Orleans, cannot contribute their indirect costs toward the matching part of the budget. And since the Louisiana State Library is funded to a great extent by federal funds, it cannot contribute anything to the matching portion. LSU-Baton Rouge will keypunch both its own LC card numbers and those of Louisiana State Library, hence its keypunching budget is substantially larger than those of the other participants.

Budget

Part I. Funds to be expended from grant.

Salaries

1.	Clerical help (at Louisiana State Library) to extract LC numbers.		$3,000.00
2.	Keypunch operators to keypunch at least 250,000 LC numbers @ $2/hr.		
	A. LSU-B.R. operators	($2123.00)	2,123.00
	B. LSU-B.R. indirect costs	(678.00)	678.00
	C. LSU-N.O. operators	(330.00)	330.00
	D. LSU-N.O. indirect costs	(154.00)	154.00
	E. USL operators	(458.00)	458.00
	F. La. Tech operators	(457.00)	457.00

Contractual Services

3.	Computer time, Spectra, Model 46, $200/hr., for at least 2 hours	400.00
4.	Printing 50 copies of 350 pp., includes cost of masters and paper	350.00

Supplies

5.	Punch card stock, to punch 250,000 numbers, 4 to a card: 32 boxes (2,000 cards/box) plus 10 extra, 42 boxes @ $1.20/box	50.00
	Total	$8,000.00

Part II. Matching funds.

Salaries

6.	Administrative and Supervisory	
	A. Project Director, 10% of time from March 1, 1971 to June 30, 1971	600.00

B. Programmer and data processing supervisor, 10% of time from March 1, 1971 to June 30, 1971 — 325.00

C. Administrators and supervisory time from

(1) LSU-Baton Rouge — 500.00

(2) La. Tech — 490.00

(3) LSU-New Orleans — 500.00

(4) USL — 600.00

D. Secretarial help, USL, 10% of time — 220.00

E. Other clerical help, all institutions estimated 50% of item 1, above, 4 libraries — 1,200.00

7. Keypunch verification at LSU-B.R., USL, LSU-N.O. and La. Tech — 3,000.00

Contractual Services

8. Keypunch rental - 4 IBM 029, at $72/mo. for 4 mo. — 1,152.00

9. Time-sharing terminal, data phone set and Bell Telephone charges — 50.00

10. Additional Computer time, Spectra, Model 46, $200/hr., 2 hours — 400.00

11. Communication, telephones, postage — 100.00

Materials, Supplies, Equipment

12. Computer printing paper — 20.00

Travel

13. In State, between libraries for project conferences, 5 persons, at least 3 trips, @ $15/trip — 125.00

Indirect Costs

14. USL. 28.6 of the following

A. Item 2E — 458.00

B. Item 3 — 400.00

C. Item 4 — 350.00

D. Item 5	50.00	
E. Item 6A	600.00	
F. Item 6B	325.00	
G. Item 6D	220.00	
H. 1/4 of Item 6E	300.00	
I. 13.3% of Item 7	400.00	
J. 1/5 of Item 8	230.00	
K. Item 10	50.00	
L. Item 11	400.00	
M. 1/5 of Items 12, 13, 14	245.00	
28.6% of	$4,028.00	1,152.00

15. LTU - 26.32% of following:

A. Item 2F	457.00	
B. Item 6C(2)	490.00	
C. 1/4 of Item 6E	300.00	
D. 13.3% of Item 7	400.00	
E. 1/5 of Item 8	230.00	
F. 1/5 of Items 12 & 14	45.00	
26.32% of	1,922.00	506.00

16. LSU-N.O. - 31.8% of following:

A. Item 6C(3)	500.00	
B. 1/4 of Item 6E	300.00	
C. 13.3% of Item 7	400.00	
D. 1/5 of Item 8	230.00	
E. 1/5 of Items 12 & 14	45.00	
31.8% of	$1,475.00	469.00

17. LSU-B.R. - 31.9% of following:

A. Item 6C(1)	500.00	
B. 1/4 of Item 6E	300.00	
C. 3/5 of Item 7	1,800.00	
D. 2/5 of Item 8	461.00	
E. 1/5 of Items 12, 14	45.00	
31.9% of	$3,106.00	991.00
Total matching		$12,400.00
Total project cost		$20,400.00
Total amount of request		$8,000.00

PROPOSAL TO CREATE A RETROSPECTIVE UNION CATALOG OF BOOKS IN PARTICIPATING LOUISIANA LIBRARIES BASED ON LIBRARY OF CONGRESS CARD NUMBERS USING THE COMPUTER TO SORT AND LIST THE NUMBERS IN PRINTED FORM

PROPOSAL TO CREATE A RETROSPECTIVE UNION CATALOG OF BOOKS IN PARTICIPATING LOUISIANA LIBRARIES BASED ON LIBRARY OF CONGRESS CARD NUMBERS USING THE COMPUTER TO SORT AND LIST THE NUMBERS IN PRINTED FORM

PHASE TWO - Data Input, Storage and Retrieval

Purpose: Phase Two of the retrospective project will consist of storing all the LC card numbers on a magnetic disk, using the University of Southwestern Louisiana's computer facilities, in such form as to allow retrieval of the individual numbers either by a paper listing or through a time-sharing remote access terminal. These terminals could be either teletypewriters or video display devices.

Part of the project will be to determine which of the two retrieval devices would be most economical or practical. Choice will depend partly on (1) the volume of requests and (2) the comparative costs of either approach. If the paper print-out is chosen it will mean printing a periodic updated list of the LC numbers. There is a demonstrated convenience in using the printed list although to produce the list is time-consuming and the list would contain a very large number of numbers which would never be consulted so much of the listing would be extraneous.

On-line terminals, on the other hand, although initially expensive and with continuing maintenance costs, can also be used for many other functions.

Methodology: Data on punch card or paper tape will be sent to the University of Southwestern Louisiana, loaded into the computer, stored directly on disks, to be sorted in format readily retrievable. Terminal time-sharing devices will be installed on the premises of selected participants so that any of the participants could call the computer directly and receive an immediate response. (Time needed to input the data once acquired should be very short. However, the time needed for feasibility testing could be as long as a year).

Those libraries without a time-sharing terminal could either telephone USL directly or send a message on existing TWX installations and USL would in turn, query the computer.

Budget

Salaries

1.	Project Director @ 10% of 12 mo.	$1,800.00
2.	Programmer-analyst @ 10% of 12 mo.	1,000.00
3.	Statistical Consultants	700.00
4.	Secretary @ 10% of 12 mo.	650.00
5.	Fringe benefits, 10% of items 1 - 4	415.00

Contractual Services

6.	Computer, Spectra 70, Model 46, $200/hr., 20 hrs.	4,000.00
7.	Teletypewriter time-sharing terminals, 6, at $85/mo., 12 months	6,120.00
8.	Printing costs, limited distribution, 50 copies	2,000.00
9.	Communication, telephones, postage	250.00

Materials, Supplies, Equipment

10.	Magnetic Disk (includes monthly maintenance)	500.00
11.	Video Display Cathode Ray Time-sharing Terminal →	8,712.00
12.	Print paper to print 1,000,000 numbers, 400 numbers/sheet: 2500 sheets per copy, 50 copies = 125,000 sheets, @ 2.60/1000 sheets	325.00
13.	Monroe 1655 desk calculator	3,700.00

Travel

14.	National, 3 trips @ 400 each	1,200.00
15.	In state, between libraries and for project conferences	2,000.00

Total		33,372.00
16.	Miscellaneous, 10% of items 1 - 15	3,337.00
17.	Total Direct Costs (sum of items 1 - 16)	36,709.00
18.	Indirect Costs, 26.32% of direct costs (item 17)	9,662.00
19.	Total Project Cost (item 17 - 18)	46,371.00
20.	Less USL's Contribution (item 18)	-9,662.00
21.	Total Proposed Budget (rounded)	36,700.00

LNR: NUMERICAL REGISTER OF BOOKS IN LOUISIANA

By
WILLIAM E. McGRATH

University of Southwestern
Louisiana Library

Reprinted from Louisiana Library
Association Bulletin

Fall, 1971
Volume 34, Number 3

A ONE MILLION VOLUME COMPUTER OUTPUT MICROFICHE NUMERICAL UNION CATALOG IN LOUISIANA, WITH A STATISTICAL SUMMARY

ABSTRACT

A union catalog of 1,100,000 books on computer output microfiche in twenty-one Louisiana libraries is described. The catalog, called LNR for Louisiana Numerical Register, consists not of bibliographic information, but primarily of the LC card order number and letter codes for the libraries holding the book. The computer programs, the data bank and output are described. The programs provide the capability for listing two million or more books. Also described are the elaborate by-product statistical tabulations which provide a rich source for analysis.

A ONE MILLION VOLUME COMPUTER OUTPUT MICROFICHE NUMERICAL UNION CATALOG IN LOUISIANA, WITH A STATISTICAL SUMMARY

WILLIAM E. MCGRATH: Director of Libraries, and DONALD SIMON: Systems Analyst and Computer Programmer, University of Southwestern Louisiana Library, Lafayette, Louisiana.

Twenty-one Louisiana libraries have produced on Computer Output Microfiche (COM) a Union Catalog containing locations for 1,100,000 books. About 150,000 of these are current acquisitions (books acquired in the last two years); the rest are volumes in the retrospective collections of ten of the twenty-one libraries. The Numerical Register of Books in Louisiana Libraries, as the Catalog is now entitled, is the second step toward what is hoped will be a comprehensive current and retrospective list of two million volumes or more, the estimated holdings of the participating libraries. The first was a conventionally printed Register of 550,000 books, issued in 1971 and distributed to 50 Louisiana libraries.

The new Register is not a bibliography. It includes no bibliographic information. It is a location device for books whose bibliographic information is already known, and includes nothing that is not also listed by the Library of Congress. The title was deliberately chosen to distinguish it from an older bibliographic Louisiana Union Catalog. All books listed in the Register are those having an LC number; indeed the LC number is the entry. The term "numerical" was chosen because we anticipate using other numbers besides the LC number--e.g., the Mansell number, and the ISBN.

The LC card order number is the nearest to a universal book number we have. This fact is put to good use by the Library of Congress in its own NUC - Register of Additional Locations. There are other LC number indexes, but they are not union catalogs. (The Mansell number, of course, will be very useful when publication of the NUC - Pre-1956 Imprints is complete).

Many more titles can be represented on a page by number codes than by complete bibliographic data, say 600 compared to 9. Unit costs are, therefore, much less. The first edition (1971) containing 550,000 volumes was produced for an estimated total cost of $20,600--$8,600 grant plus $14,000 absorbed. One hundred copies of the Register were printed in conventional form with approximate overall unit costs for keypunching, computer, travel, salaries and printing, as follows:

	In terms of grant funds only	In terms of total funds, grant plus absorbed
Per title entry	2.5¢	6.0¢
Per volume entry	1.6¢	3.8¢

The second edition (November, 1972) contains over 1,100,000 volumes and in terms of the second grant, was produced on Computer Output Microfiche for an estimated total cost of $31,200, i.e., $10,000 grant plus $21,200 absorbed. (Reproduction costs for the COM are negligible. For an original copy of 5 fiche, containing all 1,100,000 volumes, we were charged $25 by a commercial firm, and for extra copies, $3 each. Copies for distribution will be sold at a slightly higher price). Unit costs for the COM edition are

	In terms of second grant funds only	In terms of total funds, second grant plus absorbed
Per title entry	1.8¢	5.6¢
Per volume entry	.9¢	2.8¢

Unit costs computed on the basis of total costs to date, suggest that they remain relatively constant from cumulation to cumulation.

The concept of a numerical register is not new. The idea was discussed at length in a proposal by Harry Dewey (1) almost a generation ago in which

he espoused all the essential ideas and again in 1965 by Louis Schreiber (2). They argued that if the bibliographic data, including the LC number were already in hand, one could then merely look up the number in a numerical union catalog to determine a location. Goldstein and others (3) have also studied what they called the "Schreiber catalog" and have produced a sample computer printout of LC numbers. Computer output microfiche, on the other hand, was not anticipated in the original concept. It has made reproduction and distribution cheap, fast and eminently feasible. The history of the Register and its rationale have been discussed more fully by McGrath (4).

PROGRAMS COMPRISING THE UNION CATALOG SYSTEM

The Union Catalog data record is shown below. The first three fields are the familiar LC card order number.

(1)	(2)	(3)	(4)
ALPHA series	Year or numeric series	Serial number within numeric series	Library
Agr	69	2354	C

(1) Alpha series prefix--this data field may contain from 1 to 4 alphabetic characters denoting a special series.

(2) Numeric series prefix--this data field may contain 1 or 2 digits.

(3) Serial number--this data field may contain up to 6 numeric digits.

(4) Alphabetic library designation code--this field contains a preassigned alphabetic code (up to 26) designating the participating library.

The three programs comprising the Union Catalog system which use the above data record are shown in Figure 1 and described below.

LNREDT Program

LNREDT is an editing program which examines all card input data to determine whether acceptable or not.

Each data field as shown above is examined as follows:

Field 1 for the presence and rejection of non-alphabetic characters, and also to determine if the alphabetic code is a member of the accepted set of codes obtained from the Library of Congress; the accepted records are transferred to a magnetic tape file for subsequent use;

Fields 2 and 3 for the presence and rejection of non-numeric characters;

Field 4 to determine if alphabetic.

LNRSRT Program

LNRSRT is a sorting program which sorts all records on the above mentioned tape file. The major sort key is the numeric prefix, Field 2. The minor sort keys in order of the sort sequence are:

Field 1--the alphabetic special series indicator;

Field 3--the book serial number;

Field 4--the library code designation.

LNRLST Program

LNRLST is the main program which uses the sorted input data tape to:

a. create a single record for each unique LC number containing the library code designation of each library having this particular book;

b. produce a listing of the above records in chronological order;

c. enter into a memory matrix the combinations of libraries created in part (2); combinations are then counted; each time a combination is encountered, the matrix is searched for a match; if a match is found, the corresponding matrix position is incremented by one; if no match is found, a new matrix position is created with the new combination and the corresponding count initialized to one; this routine also provides for a total count of each library's contributions plus a grand total of all libraries' contributions;

d. tabulate, from the data compiled in (c) above, several elaborate tables of summary statistics; these statistics are described later in this paper.

The number of libraries the program LNRLST can accomodate is a variable and is entered as a run-time parameter along with the library names and code designations. The main program occupies approximately 150,000 bytes of core memory.

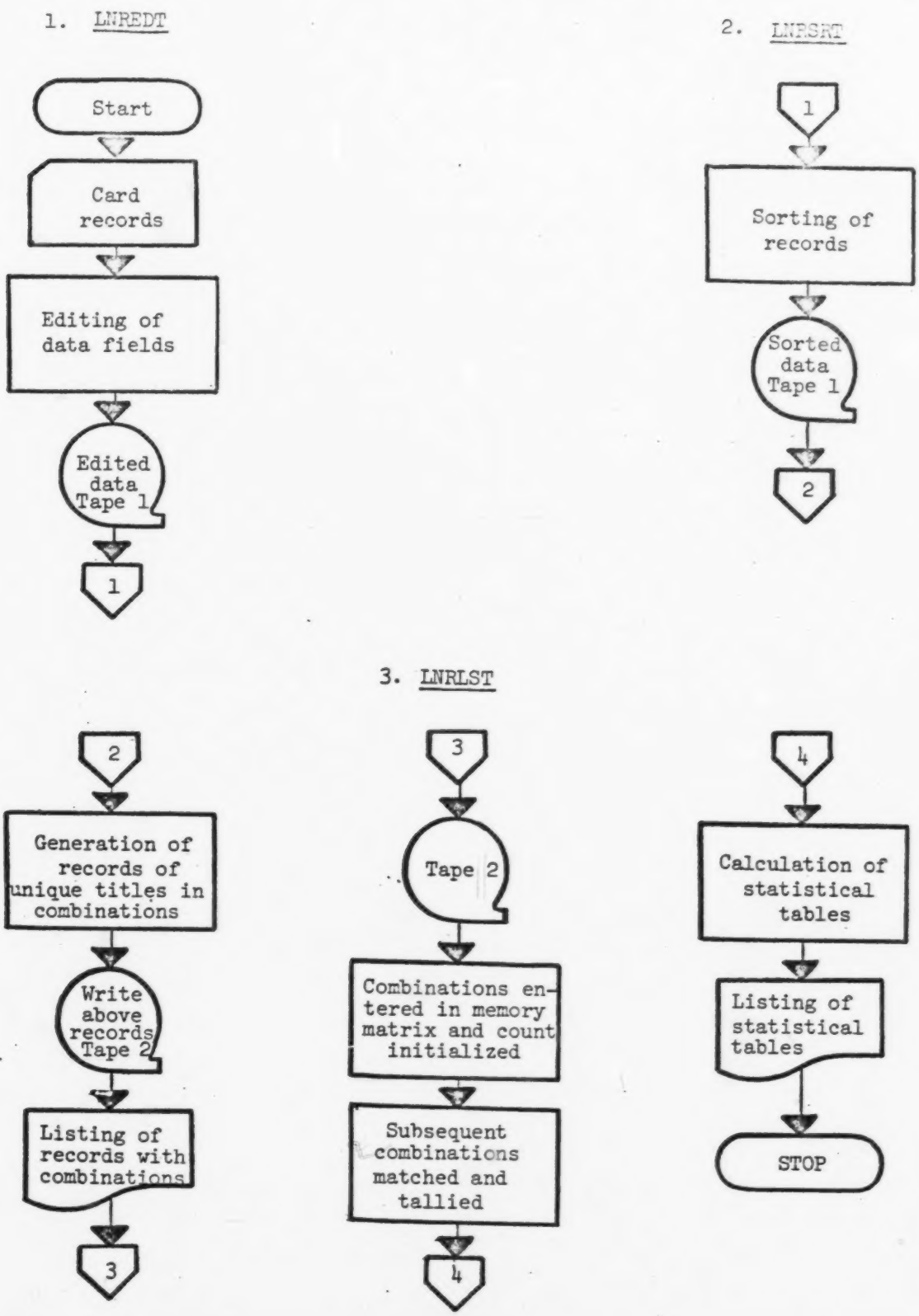

Figure 1. Flow Chart of the Programs Comprising the Register System.

THE OUTPUT

A sample of the Register entries appears in Figure 2. A simple one-letter designation was used to identify each library rather than the usual NUC designation in order to save space in the printout. These letters appear alphabetically to the right of each LC number. A typical page of the Register contains 10 columns of up to six-digit LC numbers, with the two-digit series number appearing only once at the beginning of each series. Thus each page contains about 600 LC numbers. The latest cumulation of 1,100,000 volumes (560,000 LC numbers) consists of nearly 1,000 pages. The entire output was produced on 5 pieces of fiche directly from the cumulated tape. The COM program was written by the commercial firm which contracted to run it.

The computer output microfiche was issued on five 4x6 pieces in 42X. Each piece contains 208 pages and each page contains an average of 1126 volumes and 573 titles. The data can be produced on 24X fiche as well as roll film.

76		1297	EJNP	6514	O	11161	EO	19998	C
		1447	AEZ	6652	E	11220	B	20387	E
628706	ACE	1566	C	6758	CEI	11267	H	20962	A
628714	ACE	1648	J	6896	H	11338	E	21592	E
628969	E	1672	L	6959	A	11342	Z	22250	A
629194	E	1865	A	6967	E	11417	AE	22731	E
629265	EH	1979	E	7049	P	11441	CHIJ	23416	C
629640	B	2050	C	7187	ABEH	11713	E	28290	C
629876	ABEH	2121	BC	7231	AB	12055	B	31175	C
630796	A	2235	A	7302	E	12264	E	32603	C
630911	E	2357	E	7345	E	12449	A	41981	A
631060	EH	2365	O	7408	O	12737	E	63281	C
631158	E	2381	EO	7538	E	12796	CE	75011	E
631316	E	2452	C	7570	E	12808	AE	75046	AEMOZ
631411	A	2755	A	7593	M	12816	E	75054	LZ
632084	E	3153	AE	7668	CJM	12970	MO	75097	Z
632112	E	3161	C	7712	AJ.	13316	J	75109	ABHNZ
632967	Z	3208	C	7763	E	13446	E	75117	C
633345	E	3259	E	7842	EI	13454	E	75125	EZ
633416	AE	3370	BCKO	7921	I	13584	EP	75133	AE
650787	E	3484	O	7964	EO	13626	Z	75141	E
653257	E	3539	O	8089	O	13880	A	75168	CEOZ
653466	E	3547	AE	8168	E	13927	E	75176	A
653711	F	3555	E	8176	BE	13978	E	75212	ENO
653754	E	3634	B	8263	C	14104	E	75220	O
654088	E	3705	E	8622	E	14234	AE	75255	E
750365	C	4020	E	8649	H	14277	E	75326	Z
835133	E	4256	EM	8657	E	14774	H	75334	Z
902280	E	4280	E	8760	E	14802	E	75342	AE
902552	E	4299	CO	8823	JLO	14967	EO	75377	O
908363	A	4300	E	8902	B	15073	A	75385	EO
930993	O	4378	E	8910	E	15388	E	75393	AZ
973109	O	4449	E	9000	B	15424	EHO	75448	AEZ
973606	O	4587	CE	9097	O	15440	E	75456	AZ
		4607	E	9157	AE	15503	C	75500	EZ
PS	76	4690	BCEN	9236	B	15972	O	75527	EJ
		4729	M	9314	Z	15980	E	75535	BO
15168	J	4788	E	9611	E	16003	E	75551	E
112600	J	4859	C	9717	O	16109	E	75578	M
		4891	E	9792	BE	16141	EO	75586	A
		4903	ACEO	9944	Z	16393	A	75614	E
		4911	E	10294	O	16405	E	75630	ELMO
	77	4938	E	10349	E	16472	E	75728	A
		5087	BJLO	10354	Z	16649	E	75736	Z
5	A	5158	AB	10357	E	16681	E	75779	AI
56	I	5190	A	10361	J	16728	E	75787	AE
100	O	5296	O	10365	E	16752	E	75823	AE
214	BP	5564	C	10460	E	17260	CE	75866	ABIZ
257	BE	5568	E	10468	A	17567	E	75874	EZ
360	A	5647	E	10558	E	17689	E	75902	ACEGL
407	A	5655	A	10631	Z	17733	O		Z
431	CP	5785	O	10645	E	18103	E	75937	ABCMN
553	C	5813	AE	10661	AE	18154	E		OZ
632	E	5821	BP	10716	EO	18225	E	75996	C
738	AE	5927	E	10723	Z	19038	E	76051	ACIOP
876	ABCEH	6112	E	10774	B	19056	E		Z
	L	6163	ABJ	10895	AZ	19273	E	76086	O
1002	O	6242	E	10941	E	19415	E	76094	AH
1246	I	6372	E	11031	E	19510	E	76130	Z
1270	E	6506	B	11129	E	19667	A	76157	E

Figure 2. Portion of a typical page of the computer print-out showing the 2-digit 76 and 77 series, a typical prefix--PS, the serial numbers with the series, and letter codes to the right of each serial number. For example, Library A has the book 77-5; seven libraries--A, B, C, M, N, O, and Z hold the book 77-75937.

STATISTICAL SUMMARY

The large samples of holdings (from an initial 5,000 numbers, through successive cumulations to 90,000 and, the most recent, 1,100,000) provide an excellent data base for statistical analysis. We believe the samples may be the largest title by title comparison of monographs ever tabulated in this format. Very little analysis is presented in this paper, but the data base and its format will be explained. Even without analysis, many interesting observations can be made.

Most of the tabulations are designed to throw light on the various aspects of the overlap problem, since a decisive factor in determining the utility of the Register is a knowledge of the number of titles held in common by all the libraries. Over the years there has been continuing interest in overlap. Probably the first and most elaborate of the early studies was by Leroy Merritt (5), and one of the most recent by Leonard, Maier and Dougherty (6). Continuing interest is expressed in such proclamations as that by Ellsworth Mason where he claims that materials are "being acquired in duplications that are rather staggering across the country."(7).

The following statistics were tabulated from input for current acquisitions, the most recent being the 90,302 total, rather than the retrospective and current totals in the production runs. The 90,302 volumes were acquired, for the most part, during the two year period fall, 1969 to fall, 1971. The statistics show holdings for sixteen libraries.

Table 1. Titles Held in Common by Each Unique Combination of Libraries.

%	Library Combination	Titles in Common	Combined Holdings
.48	AB	52	10688
	ABC	1	17041
	ABCE	2	46227
	ABCEGHJLZ	1	69111
	ABCEHMPZ	1	65485
	ABCEIZ	1	56754
	ABCEL	1	51118
	ABCEM	1	50030
	ABCEMZ	1	57792
	ABCHIMPZ	1	39064
	ABCHJLMP	1	37581
	ABCIJKMPZ	1	39308
	ABCIJLZ	1	36612
	ABCJNOZ	1	39215
	ABCL	1	21932
	ABCM	1	20844
	ABCU	1	23009
	ABE	2	39874
	ABEH	2	44346
	ABEHIJM	1	55067
	ABEHIJMNPZ	1	66188
	ABEHJ	1	48499
	ABEHJMZ	1	60064
	ABEHJN	1	52790
	ABEHNO	1	54117
	ABEHZ	1	52108
	ABEJOZ	1	57757
	ABEL	1	44765
	ABEMZ	1	51439
	ABEO	1	45842
	ABEZ	4	47636
	ABGHLMP	1	28681
	ABGL	1	17185
.01	ABH	3	15160
.01	ABHZ	3	22922
	ABI	1	13453
	ABJ	1	14841
	ABJKLP	1	23516
	ABKNO	1	21510
.01	ABL	3	15579
	ABLP	1	18700
	ABM	1	14491
.01	ABN	2	14979
	ABO	1	16656
	ABP	1	13909
.09	ABZ	18	18450
.30	AC	34	11061
.07	ACE	32	40247
	ACEG	1	[illegible]
	ACEGHJNO		
	AC[illegible]		

%	Library Combination	Titles in Common	Combined Holdings
.18	AHZ	32	16942
.22	AI	17	7473
	AIJ	1	11626
	AIJLNOP	1	29997
	AIJZ	1	19388
.01	AIL	2	12364
	AIMO	1	17244
	AIN	1	11764
.01	AIZ	3	15235
.46	AJ	41	8861
.01	AJL	2	13752
	AJLNO	1	24011
.02	AJM	3	12664
	AJMZ	1	20426
	AJN	1	13152
	AJNP	1	16373
	AJO	1	14829
.01	AJP	2	12082
	AJPZ	1	19844
.07	AJZ	13	16623
.03	AK	2	5271
	AKZ	1	13033
.32	AL	31	9599
.01	ALM	2	13402
.01	ALMP	2	16623
	ALMZ	2	21164
	ALN	1	13890
	ALNZ	1	21652
.02	ALZ	4	17361
.34	AM	29	8511
.01	AMN	2	12802
.01	AMO	2	14479
	AMP	1	11732
.04	AMZ	8	16273
.36	AN	33	8999
	ANO	1	14967
.04	ANZ	7	16761
.16	AO	18	10676
.02	AOZ	4	18438
.11	AP	9	7929
5.43	AZ	678	12470
1.18	BC	146	12333
.03	BCE	14	41519
	BCEFO	1	48067
	BCEG	1	43[illegible]
	BCEH		
	BC[illegible]		

%	Library Combination	Titles in Common	Combined Holdings
.07	KO	5	6531
.10	KP	4	3784
.01	KZ	1	8325
.83	LM	73	8694
.06	LMN	8	12985
	LMNP	1	16206
	LMNZ	2	20747
.04	LMO	6	14662
	LMOP	1	17883
.02	LMP	3	11915
.02	LMZ	4	16456
.90	LN	83	9182
.01	LNO	3	15150
.04	LNP	5	12403
.01	LNZ	3	16944
.70	LO	77	10859
.02	LOP	4	14080
	LOZ	1	18621
.80	LP	65	8112
	LPZ	1	15874
.37	LZ	48	12653
.65	MN	53	8094
.02	MNO	4	14062
.01	MNP	2	11315
	MNZ	1	15856
.62	MO	61	9771
.01	MOP	2	12992
.01	MOZ	3	17533
.49	MP	35	7024
	MPZ	1	14786
.29	MZ	34	11565
.43	NO	45	10259
.01	NOP	2	13480
	NOPZ	1	21242
.01	NOZ	2	18021
.69	NP	52	7512
	NPZ	1	15274
.31	NZ	38	12053
.40	OP	37	9189
	OPZ	1	16951
.29	OZ	40	13730
.18	PZ	20	10[illegible]

Acquisition and Circulation, 12/13/72

The Basic Tabulation--Titles Held in Common by Unique Combinations of Libraries

The basic tabulation sections of which are shown in Table I, actually fill seven pages of computer printout. The tabulation is designed so that each unique and actual combination of libraries is separately listed, and the books held by each combination are counted. Thus, in the Table, although the total number of books held in common by Libraries A and B are 127, the number of books held in common by them and no other library is only 52. The number of books held by Libraries A, B and Z, and no other library is 18. None of these 18 are included in the count of 52, nor none of the 52 in the 18. They are mutually exclusive. But the 18, plus the 32, plus the small counts in each of the other combinations in which A and B share holdings is 127.

The percentages of common holdings for each combination is also given, except for reasons of space, whenever the percentage is less than .01. Thus libraries A and B have .48% in common of their total combined holdings of 10,688 volumes.

Interesting to note that of the 65,535 possible combinations, in only 444 combinations did the percentage of common holdings exceed .01%, and in only 8 did the percentage exceed 1%. Of these, the highest is 5.43% (A and Z). This 5.43% means that 678 of A and Z's common holdings were held by no other library. The total of A and Z's common holdings that were also held by other libraries is 1315, or about 10.5% of 12,470. Again this is the highest percentage of any combination. All other combinations have less than 10%.

Summary of Titles Held in Common

The basic tabulation of titles held in common is summarized in Table 2. Taking the first row across the page, Column 1 is the number of libraries from 1 to 16 in each combination. Column 2 is the total number of titles counted in all combinations. For example, 59,907 titles exist in unique copy, thus there were only 59,907 copies (column 3), but there were only 8 titles which as many as 9 libraries held, for a total of 72 copies (column 3).

Column 4 shows that all 16 libraries contributed unique titles and that there were 117 different combinations of two libraries, out of a possible 120 (column 5). Thus there were three combinations of two libraries which had no titles in common. It is also most interesting that there were only seven combinations of nine libraries out of a possible 11,440, and no combinations of ten or larger.

According to the binomial distribution, there are 65,535 theoretical ways that 16 libraries can combine (total, column 5), whereas, in this sample, only 1,182 combinations occurred (total, column 4).

Column 6 is the result of column 2 divided by column 4. Thus 3774.19 is the average number of unique titles contributed by each library. 74.92 is the average number held by any combination of two libraries, and 6.89 is the average held by any combination of three.

Table 2. Summary of Titles Held in Common by Unique Combinations of Libraries (Spring, 1971 tabulation).

Column 1 No. of Libraries in Each Combination	Column 2 Total No. of Titles in all Combinations	Column 3 Total No. of Copies in all Combinations	Column 4 No. of Times a Combination Occurred	Column 5 Theoretical No. of Times a Combination can Occur (Binomial Distribution)	Column 6 Average Title Overlap Per Combination
1	59,907	59,907	16	16	3,774.19
2	8,766	17,532	117	120	74.92
3	2,453	7,359	356	560	6.89
4	782	3,128	360	1,820	2.17
5	279	1,395	214	4,368	1.30
6	84	504	75	8,008	1.12
7	43	301	41	11,440	1.04
8	13	104	12	12,870	1.08
9	8	72	7	11,440	1.14
10	0	0	0	8,008	0.00
11	0	0	0	4,368	0.00
12	0	0	0	1,820	0.00
13	0	0	0	560	0.00
14	0	0	0	120	0.00
15	0	0	0	16	0.00
16	0	0	0	1	0.00
Totals	72,335	90,302	1,182	65,535	61.19

Summary of Each Library's Multiplicated Titles

The administrators of each library are especially interested to know how many of their own titles are also held by other libraries. This information for total input (i.e., for titles with LC prefixes from 1900 to the present) is given in Table 3. (Tables were also produced giving the same kind of information by decade and for the last two years, but are not reproduced here.)

Table 3. Summary of Each Library's Multiplicated Titles (1900-1971 imprints)

Column 1 Library	Column 2 Library Code	Column 3 Number of Volumes Contributed by Each Library	Column 4 Each Library's Volume as a % of Total Volumes	Column 5 No. of Titles for Which Copies are also Held by Other Libraries	Column 6 Each Library's Multiplicated Titles as a % of Grand Total (Col.5:Col.3)	Column 7 Each Library's Multiplicated Titles as a % of Grand Total (Col.5:Total, Col.3)
Louisiana State Library	A	4,708	5.21	2,497	53.03	2.76
Louisiana Tech University	B	5,980	6.62	2,378	39.76	2.63
University of Southwestern Louisiana	C	6,353	7.03	1,932	30.41	2.13
Louisiana State University-Baton Rouge	E	29,186	32.32	6,190	21.20	6.85
Louisiana State University Medical Center	F	580	.64	168	28.96	.18
Grambling	G	1,606	1.77	471	29.32	.52
Centenary	H	4,472	4.95	2,061	46.08	2.28
Louisiana State University-Alexandria	I	2,765	3.06	1,087	39.31	1.20
Southeastern	J	4,153	4.59	1,849	44.52	2.04
Northwestern	K	563	.62	230	40.85	.25
Northeastern	L	4,891	5.41	1,980	40.48	2.19
Loyola	M	3,803	4.21	1,744	45.85	1.93
Louisiana State University-Shreveport	N	4,291	4.75	1,749	40.75	1.93
Louisiana State University-New Orleans	O	5,968	6.60	1,733	29.87	1.97
Nicholls	P	3,221	3.56	1,048	32.53	1.16
New Orleans Public	Z	7,762	8.59	3,228	41.58	3.57
Totals		90,302	100.00	30,395		
Average		5,644	6.25	1,900	37.78	2.09

The column labels are self explanatory, but it may be observed that the total in Column 5, 30,395, equals the difference between the total copies, 90,302 (Column 3, Table 2) and the number of titles held by one library only, 59,907 (Columns 2 and 3, Table 2).

Distribution of Titles Published and Multiplicated by Decade.

Table 4 shows that the very largest overlap, in current acquisitions, occurs among books with recent imprints. This is to be expected since these figures do not make any comparison to older books recently acquired by one library to those already in another library, and since the acquisition of older books is from a much larger universe than that for current books.

Table 4. Distribution of Contributed Titles Published and Multiplicated by Decade (Titles acquired from 1969 to 1971).

Imprint Period	Number of Titles Contributed	% of Titles Contributed	Number of Volumes Multiplicated	% of Total Volumes Multiplicated
1900-1909	1483	2.05	23	.13
1910-1919	1049	1.45	29	.16
1920-1929	1180	1.63	22	.12
1930-1939	1816	2.51	74	.41
1940-1949	2539	3.51	102	.57
1950-1959	5353	7.40	361	2.01
1960-1971	58915	81.41	17356	96.59
Totals	72335	100.00	17967	100.00

OTHER SUMMARY STATISTICS

The foregoing tables illustrate the kind of tabulations that can be made with this type of data. More detailed tables can be compiled, and indeed were--e.g., tables giving the percentage of books acquired for each year and each decade for each library, with 10 year totals and averages. Other possibilities would be frequency distributions and summaries for clusters of similar libraries.

This material awaits analysis. We believe it contains many insights heretofore unsuspected.

FUTURE PLANS

Since the data input is so readily updatable, plans are being made to fund the extraction and keypunching of LC numbers in the remaining retrospective collections of the participating libraries. These libraries contain an estimated total of two million volumes. Succeeding cumulations will be readily produced on COM. Most of the cost has been for extracting retrospective numbers from card catalogs. Once the remaining retrospective collections are cumulated, costs for cumulating current input will be negligible.

Any final catalog of course can never list complete holdings since each library has many titles without LC numbers. Those titles could be listed in more conventional form and since they are in a minority, the expense must be far more reasonable than it would to reproduce entire holdings in conventional form.

So far we have said nothing about other aspects of the project. In committee discussions, however, much has been said about the feasibility

of using the LC card number to access the information in other major projects such as MARC, and possibly even the data bank in the Ohio College Library Resources Center. Technically, it is feasible to print a conventional bibliographic catalog by matching up our LC numbers with titles listed in the current MARC tapes; pragmatically and economically, of course, it is another matter.

Other possibilities are to print a list of specialized holdings by accessing the subject headings on the MARC tapes, assignment of specialized acquisitions, and the gathering of information which might affect development of a joint processing center.

ACKNOWLEDGMENTS

This project was supported in part by the Library Services and Construction Act Title III funds administered by the Louisiana State Library.

The authors wish to give special thanks to Miss Sallie Farrell, Louisiana State Librarian, for her enthusiastic support and fine advice. We wish also to thank the other members of the L.L.A. Committee on the Union Catalog: Mr. Sam Dyson, Louisiana Tech University; Mrs. Jane Kleiner, Louisiana State University, Baton Rouge; Mrs. Elizabeth Roundtree, Louisiana State Library; Dr. Gerald Eberle, Louisiana State University, New Orleans; Mrs. Hester Slocum, New Orleans Public Library; Mr. Charles Miller, Louisiana State University, New Orleans; Mr. Ronald Tumey, Rapides Parish Library, Alexandria; and finally, Mr. John Richard, past President of the Louisiana Library Association, who saw the importance of the project, and who appointed the original committee.

REFERENCES

1. Dewey, Harry: "Numerical Union Catalogs," Library Quarterly (January 1948), 33-34.

2. Schreiber, Louis: "A New England Regional Catalog of Books," Bay State Librarian (January 1965), 13-15.

3. Goldstein, Samuel, et al.: Development of a Machine Form Union Catalog for the New England Library Information Network (NELINET). (Wellesley, Mass.: 1970) (U.S. Office of Education final report, Project No. 9-0404.) ED 043 367.

4. McGrath, William E.: LNR:"Numerical Register of Books in Louisiana Libraries," Louisiana Library Association Bulletin, 34 (fall, 1971), 79-86.

5. Merritt, Leroy C.: "The Administrative, Fiscal, and Quantitative Aspects of the Regional Union Catalog," Union Catalogs in the United States (Chicago, Ill.: American Library Association, 1942).

6. Leonard, Lawrence E.: Maier, Joan M.; Dougherty, Richard M.: Centralized Processing: A Feasibility Study Based on Colorado Academic Libraries (Scarecrow Press, 1969).

7. Mason, E.: "Along the Academic Way," Library Journal, 96 (May 15, 1971), 1671-1676.

PREDICTING TITLE MULTIPLICATION (OVERLAP) IN A UNION CATALOG OF SIXTEEN LOUISIANA LIBRARIES USING REGRESSION ANALYSIS

PREDICTING TITLE MULTIPLICATION (OVERLAP) IN A UNION CATALOG OF SIXTEEN LOUISIANA LIBRARIES USING REGRESSION ANALYSIS

William E. McGrath, Donald Simon and Dr. Peter Dickinson
University of Southwestern Louisiana
Lafayette, Louisiana

Introduction

Sixteen Louisiana libraries are regularly contributing a record of their current acquisitions to a union catalog compiled by a computer at the University of Southwestern Louisiana. The libraries include one public library, and the state library; the rest are academic libraries.

Objective

The union catalog is being compiled on an experimental basis to determine whether its practicality and utility is satisfactory enough to justify its continuation. A decisive factor in determining its utility is a knowledge of the number of titles held in common by all the libraries. It is a simple matter to determine the number of multiplications from the existing record by comparing the unique Library of Congress card number for each book contributed. It is much more difficult to

---

The authors wish to thank the Louisiana Library Association Committee on the Computerized Union Catalog for making this study possible. No endorsement of the findings by the Committee, nor by the Association is implied, however. The conclusions and inferences are those of the authors only.

determine what the multiplication would be several years from now when presumably the file would be much larger. The question is important because the participants would like to know whether the catalog offers them any more than their own catalogs. If multiplication were high, then the catalog would offer very little to the participants except perhaps a way to distribute demand over all the libraries and to take demand load off of the large libraries. If multiplication were low, then there would be many more unique titles in the system than any one library could possibly acquire by itself.

Several prior studies have attacked the problem of multiplication[1,2,3]. (Most authors have used the term "duplication", rarely distinguishing between duplication, triplication, quadruplication, etc.) But none, so far as we know, have tried to determine the precise mathematical or statistical relationship between total input and multiplications. Indeed it is often remarked that multiplication is exponential but studies to support this vague statement are elusive.

The question of predicting multiplication can be approached in several ways. One promising approach would be to determine the functional or statistical distribution of unique titles, duplicated titles, triplicated titles, etc., among all the libraries. If this distribution could be determined, then it would hold true, no matter how many titles were input, nor how many libraries participated. More important, the distribution would presumably hold true for any group of libraries. The function for the known distribution could then be used to predict multiplication in many situations. However, this interesting question is being

left for another paper.

In this paper, only the linear and polynomial aspects of total multiplications will be tested. The statistical model for the relationships is represented by

$$Y_i = \alpha + \beta_1 X_i + \beta_2 X_i^2 + \ldots + \beta_{\kappa-1} X_i^{\kappa-1} + \varepsilon$$

where Y is the observed value, X_i is the factor that effects the response and ε is random error. This model would enable us to predict the number of future multiplications from the trend demonstrated by the existing data without a knowledge of the functional distribution. We are aware of the hazards in predicting future values from a linear or polynomial equation, hence our results will be expressed cautiously in terms of confidence limits.

Experimental Procedure

Data consisted of a record for each book with an LC card number currently cataloged (i.e., acquired) by each participant. Over a period of one year, nearly 60,000 of these numbers were submitted to the University of Southwestern Louisiana Library according to a specified standard format--i.e., the LC card number for each volume, with a code letter for the participating library. Thus,

69-3472 A

69-3472 C

69-3472 Z

identify the same book held by three libraries, A, C and Z.

The sample consists of all books acquired by 16 libraries after September, 1969; however, many of the books listed were actually published prior to 1969. At this time we have no way of knowing whether any of the libraries already owned pre-1969 imprints which appear in this sample. However, 44% of the 1968/70 books in our sample were multiplicated, whereas 13.6% of the 1900/67 books in the sample were multiplicated. Another source of possible but slight bias would be where two or more editions of the same title, each with a different LC number, were included in the sample. Whether or not such editions constitute multiplications is a matter of opinion.

The study analyzes the summary statistics from six substantial cumulations printed by USL's RCA Spectra Computer over a period of one year. The summary statistics were actually a by-product of the principle LC card number listings. The summary statistics include the total number of volumes contributed by the sixteen participants, and the total number of volumes multiplicated--that is, the total number of copies as indicated by the number of libraries holding copies. Thus if twelve volumes were contributed by three libraries, and five were held by no more than one library, the rest would have to be titles with extra copies as in the following:

	Total Titles	Total Copies	
Held by 1 library	5	5	
Held by 2 libraries	2	4	} multiplications
Held by 3 libraries	1	3	
Total	8	12	

Thus there are 8 minus 5 = 3 titles with extra copies and 12 minus 5 = 7 multiplications. These extra copies are what are referred to in this paper as "multiplications." The independent variable, X, is the number of volumes put in, and the dependent variable, Y, is the total number of multiplications. In the example above X and Y are 12 and 7 respectively.

The actual data for the six accumulations during the year are listed in Table 1.

Table 1. Statistical data summary

Accumulations	Number of Volumes (X)	Number of Multiplications (Y)
1	15,785	2,123
2	26,076	4,914
3	32,035	6,729
4	44,172	12,179
5	50,109	14,799
6	57,630	17,786

Test for a Linear Relationship

The data was submitted to a fit, by the least squares method, of the linear equation.

$$\hat{Y} = a + bX$$

and the analysis of variance is summarized in Table 2. The fitted equation is

$$\hat{Y} = -4777 + .386X.$$

Table 2. ANOVA for linear regression.

Source	df	Sum of Squares	Mean Squares	F
Due to regression	1	$1.849 \cdot 10^8$	$1.849 \cdot 10^8$	435.06*
Residual	4	$1.700 \cdot 10^6$	$4.250 \cdot 10^5$	
Total	5	$1.866 \cdot 10^8$		

*Value of F at $\alpha_{.0005}$ for 1,4 df to reject H_o:β=0 is 106

The analysis of variance shows that we can substantially reject the hypothesis (β=0) that there is no linear relationship between the number of volumes put in and the number multiplicated. We could safely assume that there is a linear relationship and for all intents and purposes we could use the linear equation to predict the number of multiplications so long as we expressed our predictions within confidence limits and so long as the size of the catalog does not outgrow the usefulness of the confidence. The confidence limits are shown in the graph in Figure 1. The graph shows that the data points through which the predicted lines has been drawn, are contained within the two outer lines of the confidence belt. As the predicted line approaches the outer reaches, the confidence belt becomes wider and thus any prediction we make for future multiplications becomes less accurate.

The graph also shows a very slow but obvious curvature of the actual data. As said before, we would be perfectly justified in making predic-

tions from the linear form if we are satisfied with the accuracy dictated by the confidence limits and with the knowledge that prediction of future values will have much less confidence. If we are interested in greater accuracy, however, the slight curvature suggests that we extend our analysis to a test of a higher order--for example, the quadratic or cubic forms of regression.

Polynomial Test

The standard polynomial computer program designated BMD05R, in BMD Biomedical Computer Programs[4] was used for this part of the investigation. This program calculates the sum of squares, the mean square, the F ratio and other statistics, due to regression of the dependent variable for each of successively higher orders of the independent variable. In addition to the standard ANOVA tables for each order, the program prints a final ANOVA table consisting of the sum of squares for each term--linear quadratic, cubic, etc. For this test, third order was deemed sufficient, so no higher order was computed. The analysis of variance for the three degrees are given in Table 3.

Table 3. Analysis of variance for 3 degree polynomial regression

Source	df	Sum of Squares	Mean Squares	F
Linear Term	1	184,959,993.4	184,959,993.4	5,162.649
Quadratic Term	1	1,230,560.00	1,230,560.00	34.35*
Cubic Term	1	391,808.00	391,808.00	10.94
Residual	2	71,653.00	35,826.562	
Total	5	186,654,014.40		

*Value of F at $\alpha_{.05}$ for 1,2 df to reject $H_0: \beta_3 = 0$ is 18.5

We see at once that the quadratic term is significant. The relationship between the total volumes put in and titles multiplicated is not linear, nor is it cubic. It is quadratic. Here, the fitted equation is

$$\hat{Y} = -1708 + .1896X + .0000026X^2$$

Referring again to the graph in Figure 1, and to the very slight trend to curve upward, we see that titles multiplicated are increasing at an increasing rate--albeit a slow rate. The curving trend, then, is slight but significant.

However, there is not too much justification for using the quadratic term since the linear term accounts for most of the variation as we can see when we compute the coefficient of determination, R^2 which is

$$\frac{1.849 \cdot 10^8}{1.866 \cdot 10^8} = .9908$$

or 99%.

Conclusion

Since the quadratic term is only slightly significant (i.e., compared to the linear term) it is as we said before, a matter of choice whether we use the linear or quadratic equation for prediction. Table 4A gives both linear and quadratic predictions for each of the given values of input volumes. The quadratic predictions, however, are undoubtedly better than the linear, as shown by the smaller residual values in the quadratic column.

Table 4A. Predicted multiplication for given input

Actual		Linear		Quadratic*	
Input	Multiplication	Predicted	Residual	Predicted	Residual
15,785	2,123	1,318.2	805	2,155	-32
26,076	4,914	5,291.8	-378	4,739	175
32,035	6,729	7,592.8	-864	6,921	-192
44,172	12,179	12,279.3	-100	12,145	34
50,109	14,799	14,571.7	227	14,761	38
57,630	17,786	17,475.8	310	17,809	-23

*From University of California BMD Biomedical Computer Programs, "BMD05R, Polynomial Regression."

Table 4B gives the linear predictions for extrapolated, or future values of input volumes, with confidence limits using the equation

$$\hat{Y}\ (\text{low}),\ \hat{Y}\ (\text{high}) = Y \pm t_{(1-\alpha/2)} S_Y \left(1+\frac{1}{n} + \frac{(X'-\overline{X})^2}{S_{XX}}\right)^{\frac{1}{2}}$$

Table 4B. Predicted multiplicated volumes for future extrapolated input.

Input Volumes	Predicted Multiplication	95% Confidence Interval for Multiplicated Volumes
75,000	24,200	21,500-- 27,000
100,000	33,800	30,000-- 37,500
125,000	43,500	38,600-- 48,400
150,000	53,100	47,000-- 59,200
200,000	72,500	63,900-- 81,100
500,000	188,000	164,000--212,000
1,000,000	381,000	331,000--431,000

The predicted values for extrapolated or future input should, as we said before, always be expressed in terms of confidence limits, and at the very least, rounded off. Furthermore, since we really could not know how the curve would behave in the future, we should have little justification in extrapolating. For example, if the pattern of con-

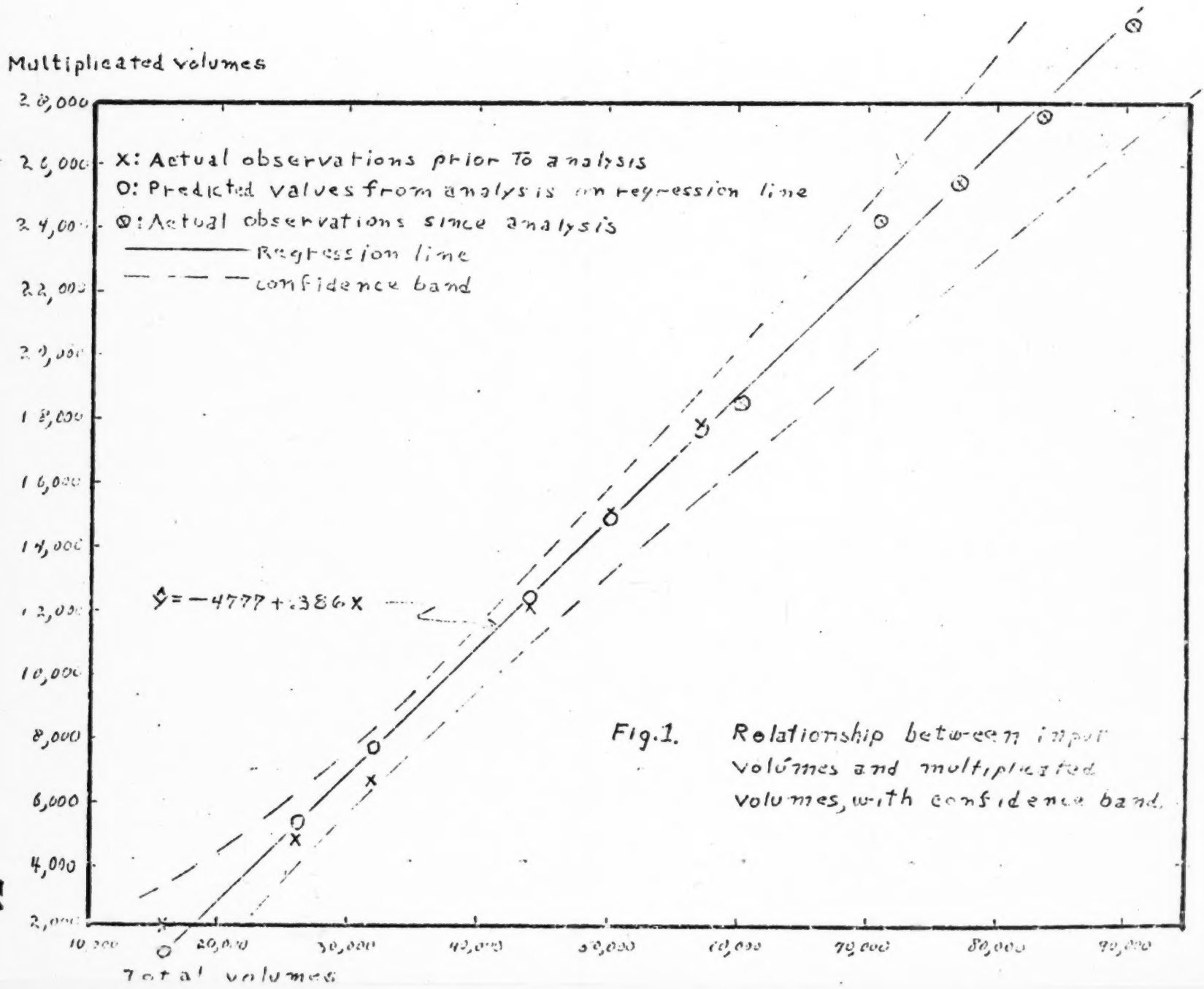
Multiplicated volumes
X: Actual observations prior to analysis
O: Predicted values from analysis on regression line
⊗: Actual observations since analysis
Regression line
confidence band
ŷ = −4777 + .386x
Total volumes

Fig. 1. Relationship between input volumes and multiplicated volumes, with confidence band.

9/9/71 McGrath

tributions changed for future input, actual multiplication could very well change accordingly.

In regression, the greater the number of data points, the greater the confidence. Six data points give minimal confidence, but we hoped that even this would prove practical in planning for future runs of the Union Catalog. Since the original analysis five more data points were added to the first six, making it possible to determine how good the first six were as predictors. The five additional points are shown in Table 5A, along with their confidence intervals calculated from the original six.

Table 5A. Later data compared to their predicted values from original equation and new equation.

Accumulations	Number of Volumes(X)	Number of Multiplications(Y)	Predicted from Original Equation	Predicted from New Equation	95% Confidence Interval
7	60,026	18,461	18,393	18,692	16,200--20,500
8	70,763	24,222	22,537	22,934	20,100--25,000
9	77,012	25,385	24,950	25,402	22,200--28,000
10	83,210	27,514	27,342	27,850	24,400--30,600
11	90,302	30,395	30,080	30,651	27,000--34,800

The new fitted equation calculated from all eleven data points is

$$\hat{Y} = -5018 + .395X,$$

very similar to the original equation. The predicted values from the new

equation are close to those from the original and well within the confidence intervals of the originals.

The analysis of variance for the extended line, calculated from the eleven data points is shown in Table 5B. The linear test is still highly significant.

Table 5B. ANOVA for linear regression, with extended data.

Source	df	Sum of Squares	Mean Squares	F
Due to regression	1	$9.1186 \cdot 10^8$	$9.1186 \cdot 10^8$	2196.0*
Residual	9	$.0374 \cdot 10^8$	$.00415 \cdot 10^8$	
Total	10	9.1559		

*Value of F at $\alpha_{.0005}$ for 1,9 df to reject H_o:β=0 is 28.0

It seems clear that the relationship between multiplicated volumes and title input, for current acquisitions of the 16 libraries is similar as the total input becomes more massive.

Future Analysis Needed

Since 66% of the total volumes in the latest cumulation exist in single copies, then whatever curvilinear relationship there is must be accounted for by the remaining 34% of the data--or, the duplications, triplications, etc. It remains to be shown then, whether indeed the

multiplications, as opposed to single copies, account for the quadratic or any other effect.

In addition, as we said in the statement of Objectives, there is the very interesting question of the distribution of multiplicated titles. We have reason to believe that such a distribution may be functional.

References

1. Leonard, Lawrence E., Maier, Joan M., and Dougherty, Richard M. Centralized Book Processing, Scarecrow Press, Metuchen, N.J., 1969.

2. Dawson, John M. "The acquisitions and cataloging of research libraries: a study of the possibilities for centralized processing." Library Quarterly, v. 27, Jan. 1957, p. 1-22.

3. Nugent, William R. "Statistics of collection overlap at the libraries of the six New England state universities." Library Resources and Technical Services, v. 12, No. 1, Winter, 1968.

4. Dixon, W. J., ed., BMD. Biomedical Computer Programs. University of California Press, Berkeley, 1967.